САНКТ-ПЕТЕРБУРГСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ
КАФЕДРА КОМПЬЮТЕРНЫХ ТЕХНОЛОГИЙ И СИСТЕМ

Малютин Алексей Андреевич

Выпускная квалификационная работа бакалавра

Искусственный интеллект для обеспечения качества при подборе персонала

Направление 010300

Фундаментальные информатика и информационные технологии

Научный руководитель,
к.ф.-м.н.,
доцент
Погожев С. В.

Санкт-Петербург
2017

SAINT PETERSBURG STATE UNIVERSITY

DEPARTMENT OF COMPUTER APPLICATIONS AND SYSTEMS

Maliutin Aleksei

Bachelor's Thesis

Artificial Intelligence for Quality Assurance in Human Resource

Field of study 010300

Fundamental Informatics and Information Technology

Scientific supervisor,

Ph.D.,

Associate Professor

Pogozhev S. V.

Saint Petersburg

2017

Оглавление

Введение

Если бы я вернулся в 2016-й, я бы взял любую идею и добавил машинное обучение.

— Алекс Фламант

Много десятилетий назад писатели-фантасты предвидели будущее, где космический туризм станет обыденностью, путешествие в золотой век русской литературы — привычной альтернативой театру, а роботы и искусственный интеллект займут важнейшую роль в нашей жизни. Однако мы, как современники XXI века, можем уверенно сказать, что некоторые предвосхищения прошлого так и остались на бумаге: одни к счастью, другие к сожалению. Среди наиболее значимых технологий, перешагнувших рубеж книжной страницы, безусловно, выделяется искусственный интеллект (ИИ).

Об искусственном интеллекте ещё в 40-е годы прошлого века писал в своих рассказах Айзек Азимов, и чуть позже польский писатель-фантаст Станислав Лем. В светлое будущее ИИ верили многие: от нобелевского лауреата по экономике 1978 года Герберта Саймона[1], до американского учёного Рэйа Курцвейла, который сегодня стоит во главе программ по разработке ИИ в Google.

И хотя современный ИИ выглядит далеко не так, как люди себе его представляют, но, что более важно, уже сегодня он в той или иной степени применяется во множестве окружающих нас вещей. К примеру, именно технологии машинного обучения обеспечивают идентификацию физической активности в столь популярных сейчас фитнес-трекерах, и именно машинное обучение позволяет *очеловечить* перевод текста с одного языка на другой.

При этом, как отмечают специалисты, на рынке ещё много свободных ниш, где ИИ либо развит слабо, либо же не развит вовсе. Одной из таких сфер является сфера HR, одному из вариантов применения методов машинного

[1] Герберт Саймон (1916 — 2001) — американский учёный в области социальных, политических и экономических наук, один из разработчиков гипотезы Ньюэлла — Саймона. Книги «Модели человека» и «Наука об искусственном» оказали значительное влияние на развитие сферы ИИ.

обучения в которой и посвящена данная работа.

Сложно не согласиться с утверждением, что самым ценным ресурсом XXI века является не нефть или газ, ценнейший ресурс планеты Земля сегодня — человек с его талантами и знаниями. Если ранее существовала всем известная «золотая лихорадка», когда среди миллиардов песчинок старатели изо дня в день надеялись отыскать заветный 79-й элемент таблицы Менделеева, то сегодня такими старателями являются HR-агентства, а песчинками — миллионы людей, среди которых как раз и необходимо найти кандидата, который идеально соответствовал бы своими профессиональными и личностными компетенциями требованиями заказчика. С одной стороны, в эпоху глобализации и повсеместного распространения сети Интернет объём информации, которую необходимо обработать HR-агентству, постоянно растёт, а её структура усложняется. В то же самое время, конкуренция за таланты еще никогда не была столь жестокой: крупные компании выделяют огромные ресурсы на усиление рекрутинга, так как промедление с ответом соискателю на несколько часов может стать решающим — Ваш потенциально идеальный сотрудник уйдёт к конкуренту. Учитывая эти факты, становится очевидным, почему применение традиционных методов при работе с кандидатами – анкетирование, собеседование по телефону, личное собеседование, профессиональное тестирование соискателя и др. – имеет существенные недостатков, среди которых самым серьёзным является фактор времени.

Заключительным штрихом в описании сложившейся на рынке HR ситуации является тот факт, что сейчас сотрудники менее лояльны к своему работодателю и намного чаще готовы его сменить нежели раньше [7], люди чувствую себя более свободными как при выборе новой работы, так и в контексте ухода со старой: «Война за таланты завершилась, и таланты победили» [22]. Безусловно, данная “свобода” побуждает компании искать способы не только найма новых сотрудников, но и сохранения существующих. Одним из потенциальных способов решения этой проблемы является более

глубокий анализ информации о соискателях: не только с точки зрения соответствия *in situ*, но и с точки зрения перспективы.

Тенденцию на своеобразное перерождение сферы HR замечают как аналитики рынка, к примеру, журнал Forbes в одной из своих статей предсказывает скорый рассвет ИИ: “Основным трендом в HR станет использование машинного обучения и технологий искусственного интеллекта для подбора и оценки качеств сотрудников” [18], так и мастодонты HR, среди которых Дэйв Ульрих[2], определивший в своём интервью [13] современные технологии как один из важнейших факторов развития для отрасли в целом. Расходы на технологии не первый год показывают рост, а инвестиционный климат в сфере крайне благоприятный, что говорит об актуальности интеграции искусственного интеллекта в эту отрасль. Среди основных векторов развития Forbes особо подчёркивает те, которые связаны либо с явным применением машинного обучения, либо же частично опирающиеся на него. К примеру, это автоматизированное совмещение профиля вакансии и резюме кандидата, где предполагается использование как тривиальных агрегаторов и синтаксических парсеров, так и *supervised* методов для получения дополнительной информации о специалисте из его профилей в социальных сетях. Другой перспективной технологией, объединяющей математику и психологию, является автоматическая типологизация личности, которая должна будет заменить привычное сегодня психологическое тестирование [18].

С исследовательской точки зрения, задача применения методов машинного обучения в HR-сфере тоже выглядит крайне интересной. Прежде всего по причине того, что эта сфера очень близка к психологии, и именно в психологии горизонты развития ML ещё не достигнуты (об этом говорит, к примеру, то, что для оценки моделей используются более низкие стандарты,

[2] Дэйвид Олсон Ульрих — профессор бизнес-школы Мичиганского университета Энн Арбора, автор программ обучения с тематикой — менеджмент, управление человеческими ресурсами, а также автор 15 книг о HR и управлению.

нежели в других сферах [12]). Другой причиной является разрозненность и сильная неоднородность данных, они требуют серьёзной предварительной обработки до их интеграции в модель, поэтому хорошо реализованная предобработка данных в этом контексте может быть без особых усилий расширена и для других целей.

Если на западном рынке применение современных технологий уже как несколько лет вышло за пределы Кремниевой долины, и попытки применения ИИ в HR тем или иным образом носят систематический характер как у новичков рынка [9, 11], так и у крупнейших компаний [1, 5], то об отечественном рынке сказать аналогичное затруднительно. Уже несколько лет на российском рынке представлена Kenexa, SaaS от IBM, предлагающая возможности высокотехнологичного найма сотрудников, однако порог входа по цене крайне высок [5]. Вероятно, такие российские технологичные гиганты как Яндекс или VK тоже применяют искусственный интеллект для поиска лучших кандидатов, но это не афишируется и остаётся скрытым от глаз рынка в целом. Поэтому предложение работодателя, выразившего живую заинтересованность этим вопросом и готового к совместному преодолению трудностей, было воспринято крайне позитивно и послужило отправной точкой при написании данной работы.

Постановка задачи

Цель данной работы заключается в исследовании применимости методов машинного обучения, основанных на прецедентах, в контексте кадровой службы на рынке временного персонала. Реализация этого исследования требует всестороннего анализа исходных данных, сравнения различных по своей природе методов и дальнейшее построение лучшей модели с подбором оптимальных гиперпараметров.

Ключевой особенностью работы является то, что HR значительно ближе к психологии, и идеального результата может не быть, однако работодателем поставлена задача максимально к нему приблизиться.

Используя личную информацию о сотрудниках и результаты проведенных ими смен как обучающее множество, требуется реализовать модель, которая по новым входным данным – информация о потенциальном сотруднике – будет делать прогноз его профессиональной пригодности.

Подзадачи, последовательное выполнение которых должно позволить достичь озвученной цели, сформулированы ниже:

- Изучение предметной области:
  - Рассмотреть существующие варианты применения машинного обучения в сфере HR;
  - Определить критерии качества для сотрудника;
  - Определить наиболее перспективные к применению в данном конкретном случае методы машинного обучения;
- Анализ и предобработка исходных данных:
  - Обработка пропусков;
  - Модификация и получение новых признаков;
- Построение различных моделей на основе полученного на предыдущем этапе *dataset*-а и их сравнительный анализ;
- Верификация модели и оценка вариантов её дальнейшей интеграции в деятельность работодателя.

Среди требований, выдвигаемых к ожидаемому решению, отмечаются следующие:

- Обеспечение баланса точности и полноты предсказания, достаточных для принятия взвешенного решения менеджером отдела HR;
- Устойчивость к внешним возмущениям в исходных данных, то есть обработка различных аномалий и пиковых значений;
- По возможности минимизация времени как на построение модели, так и на осуществление предсказания;
- Возможность адаптации полученного решения к применению в смежных сферах.

Обзор литературы

Различные варианты применения машинного обучения в HR-сфере были рассмотрены во многих исследованиях, однако стоит отметить, что для их понимания требуется определённый уровень знаний.

В зависимости от вариантов использования и сферы приложения интеграция ИИ в HR сводится к следующим категориям:

1. *Автоматизация поиска кандидата под определённую позицию в компании:*

В качестве входных данных выступают полноценные резюме, поэтому применимость методов из этой категории ограничена поиском высококлассных специалистов и руководящего звена.

Статья [25] частично обосновывает необходимость ухода от ручного сравнения информации в резюме с требованиями открытой вакансии, а также содержит унифицированную методику оценивания кандидата, путём введения независимых критериев оценки резюме претендента. Актуальным в данном контексте остается и применение аппарата нечётких логик, что продемонстрировано в [24]. Работа [21] реализует идеи автоматической классификации резюме по области деятельности и их последующего ранжирования; аналогичную направленность имеет и статья [23], в которой рассматривается автоматическое извлечение информации из резюме и его последующая классификация. Среди достоинств данного нововведения подчёркивается объективизм ИИ, в частности уход от гендерной, расовой и этнической дискриминации. Согласно исследованию [8], введение *e*-рекрутинга снизило почти в 2 раза как затраты на поиск, так и среднее время закрытия вакансии.

[4] в отличие от предыдущих получает информацию напрямую с LinkedIn, и, хотя сами алгоритмы классификации и соотнесения схожи с рассмотренными ранее, в качестве дополнительно критерия выступает

результат LIWC[3] анализа, что обеспечивает учёт характера кандидата в контексте вакансии. Автор статьи [15] наглядно демонстрирует этапы построения автоматизированной системы, базирующейся на концепциях NLP для формализация текстовой информации из резюме и использующей ML для ранжирования.

2. *Предсказание результатов работы для временного персонала:*

Отличие этой категории от рассмотренной выше в ориентированности на персонал, текучесть которого крайне высока, резюме часто отсутствует, а производимые работы носят однотипный характер.

В работе [20] автор использует нейронную сеть с двумя скрытыми слоями для бинарной оценки кандидата, опираясь на такие формализованные признаки из резюме как пол, возраст, стаж работника, знание английского языка и другие. Однако, вызывает вопросы утверждение автора о «верной классификации всех работников в выборке», что может свидетельствовать о переобучении модели. Основой диссертации [19] является сравнение различных по своей природе моделей: модели бинарного выбора, применяемой в эконометрике, и нейронной сети, где на чистых численных данных (пол, возраст, наличие высшего образования, стаж, уровень владения компьютером и других, всего 11 признаков) заявлена точность модели равная 0.8.

В [6] обозначены перспективные, но пока ещё малоисследованные сферы приложения машинного обучения в HR:

- Оценка вероятности увольнения сотрудника в краткосрочной перспективе.
- Адаптация тренингов для персонала под потребности компании «здесь и сейчас».
- Автоматизация психологического тестирования кандидата.
- Автоматизации внутрикорпоративного взаимодействия между сотрудниками, анализ их потребностей и агрегация личных целей.

[3] Linguistic Inquiry and Word Count

Глава I. Предметная область

1.1 Основные понятия

Определение 1.1

Временный персонал — это персонал, нанимаемый под определенный проект со сроком найма от 6 часов до 6 и более месяцев, в зависимости от длительности проекта или обязанностей сотрудника [27].

Согласно данным аналитиков, временный персонал наиболее востребован в таких направлениях, как розничная торговля, склады, логистика, однако, в силу своей экономической привлекательности, не ограничивается только ими. К явным преимуществам данного подхода при закрытии позиции можно отнести [10]:

- Снижение затрат на содержание сотрудника в штате компании
- Облегчённое налогообложение
- Оптимизация нагрузки на HR-отдел

Услуги по аренде временного персонала чаще всего предоставляют специализированные агентства, которые стремятся максимизировать качество выполняемых работ и минимизировать свои издержки, что частично возможно, как раз путём предсказания качества сотрудника – основная задача данной работы – на этапе его внесения в базу.

В зависимости от характера и продолжительности требуемых работ выделяют две формы взаимодействия:

Определение 1.2

Аутстаффинг (англ. *outstaffing*) — это использование «внешнего» или «заемного» персонала, т.е. принадлежащего внешней организации, для решения проблемы кадрового обеспечения и интеграции интеллектуального потенциала. Данную услугу оказывают специализированные *staffing* агентства [16].

Изначально услугами аутстаффинга в России пользовались иностранные

компании при открытии локальных офисов, однако сегодня среди основных потребителей выступают компании с сезонным характером ведения бизнеса и компании, планирующие краткосрочные проекты.

ОПРЕДЕЛЕНИЕ 1.3

Аутсорсинг (англ. *outsourcing*) — передача организацией, на основании договора, определённых видов или функций производственной предпринимательской деятельности другой компании, действующей в нужной области [17].

Чаще всего на аутсорсинг передаются функция бухгалтерского учёта, поддержка IT инфраструктуры, реклама, перевод и локализация программного обеспечения, HR.

1.2 Текущая ситуация на рынке временного персонала

В сложившейся сегодня ситуации экономического спада все крупные компании тем или иным образом пытаются сократить производственные издержки, и одним из вариантов их снижения является вынесение части сотрудников за штат или привлечение временного персонала.

Если говорить о конкретных цифрах, то безусловным лидером по привлечению временного персонала является рынок розничной торговли – до 20% по России в целом и до 50% в Москве, далее следуют промышленное производство и производство товаров с высокой оборачиваемостью. Как следствие, рынок *staffing* агентств бурно развивается, и по некоторым оценкам уже близок к точке перенасыщения, поэтому с экстенсивного пути роста ведущие игроки рынка переходят на интенсивный.

В соответствии со спецификой и общим характером деятельности, рынок временного персонала обладает крайне большой текучестью, и с этим связано основное требование к *staffing* агентствам — постоянное обновление базы проверенных сотрудников, при этом требуется соблюдать баланс между скоростью обновления и качеством новых работников. В настоящий момент эту работу выполняют либо линейные менеджеры, либо специализированный

HR-отдел: получая звонок от потенциального сотрудника его персональные данные вносятся в стек *«необходима проверка»*, далее начинается итеративный процесс уточнения некоторых деталей о человеке, сверка сведений о нём с информацией из различных баз (как открытых, так и закрытых), по итогам финального интервью потенциальный сотрудник может перейти в группу *«готов работать»*. Однако, далее, после нескольких проведенных смен, возможно три варианта развития событий:

- новый сотрудник удовлетворяет фактическим требованиям агентства;
- новый сотрудник не удовлетворяет фактическим требованиям агентства и его следует удалить из базы;
- новый сотрудник несколько раз отказывался от смен, что влечёт его удаление из базы,

при этом только лишь первая ситуация позволяет агентству в будущем возместить издержки за первичную проверку, а две другие — нет.

Из описанной выше ситуации вытекает очевидное желание агентств определять *качество* потенциального сотрудника ещё на этапе внесения его в базу, тем самым снижая общие издержки на последующую проверку.

Глава II. Данные

2.1 Анализ

Исходные данные получены из Microsoft Dynamics CRM и представляют собой Excel файлы: файл с информацией о сотрудниках, принадлежащих *staffing* агентству, и файл с информацией о проведенных рабочих сменах. Связующим ключом является столбец *'ID'*, уникальный для каждого сотрудника. Суммарный объём информации: 38973 и 78197 экземпляров, соответственно; данные предоставлены за 2014–2016 года.

Признаки объекта «Сотрудник», перечислены ниже:

Признак	Тип	Количество	Unique	Top
ID (автономер в базе)		38937		
Фамилия	-	38937	20759	Иванова
Имя	Номинальный	38937	1904	Александр
Отчество	Номинальный	38937	3587	Александровна
Пол	Бинарный	38937	2	Женский
Город	Номинальный	38937	273	Москва
Гражданство	Номинальный	38937	3	Россия
Источник	Номинальный	38937	39	авито
Есть основная работа	Бинарный	38937	-	Нет
Мобильный телефон	-	38937	2	-
Дата рождения	-	38647	11356	11.01.1996
Возраст	Количественный	38635	-	-
Субъект федерации	Номинальный	36582	69	МО МСК
E-mail	Номинальный	15363	217	mail.ru

Таблица 1. Признаки

Наибольшую ценность и одновременно сложность представляют такие номинальные признаки как Имя и Отчество, имеющие большое число уникальных значений, а также Дата рождения, который в явном виде малоинформативен, однако служит хорошей базой для получения новых

признаков [3].

Рассмотрим подробнее один из номинальных признаков, к примеру, Имя. Согласно статистике, различных имён в исходных данных 1904, при этом ни один из методов машинного обучения не может явно работать со строковыми признаками: требуется конвертация строкового значения в *numeric*.

Проанализируем две крайности: самые популярные имена, встречающиеся в данных, и, как альтернативу, зависимость числа имён от числа людей с такими именами.

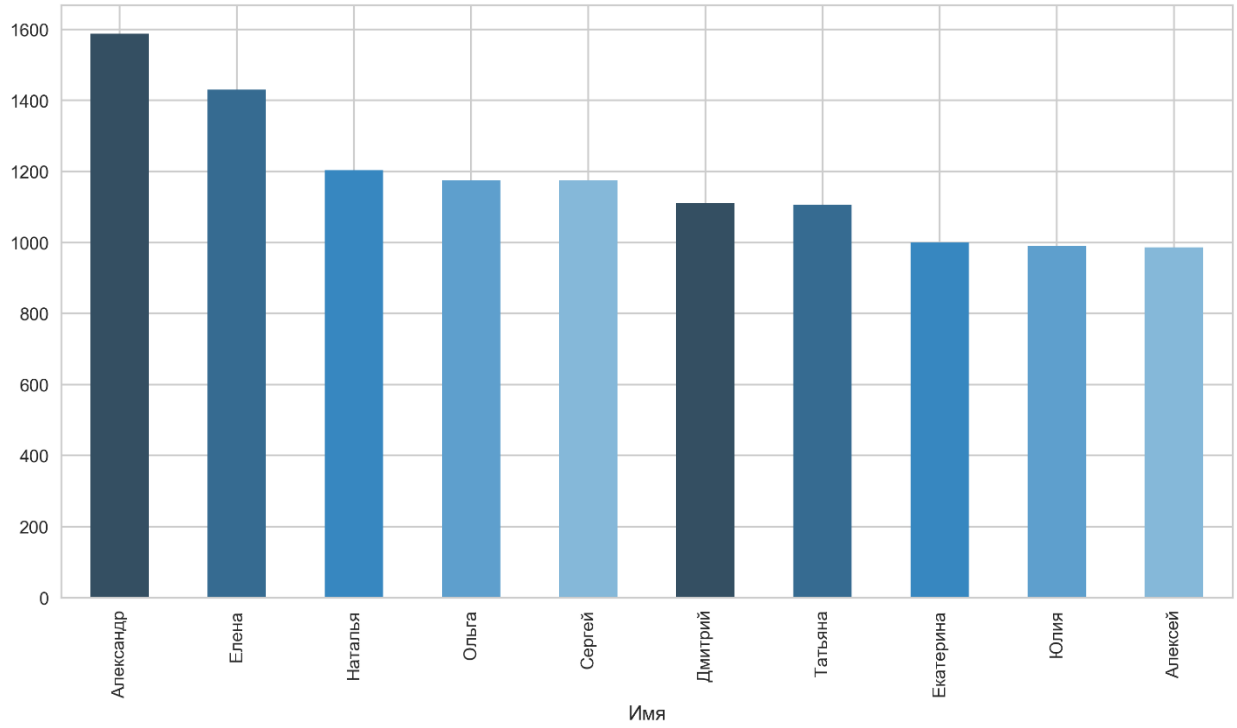

Рис. 1 Наиболее популярные имена

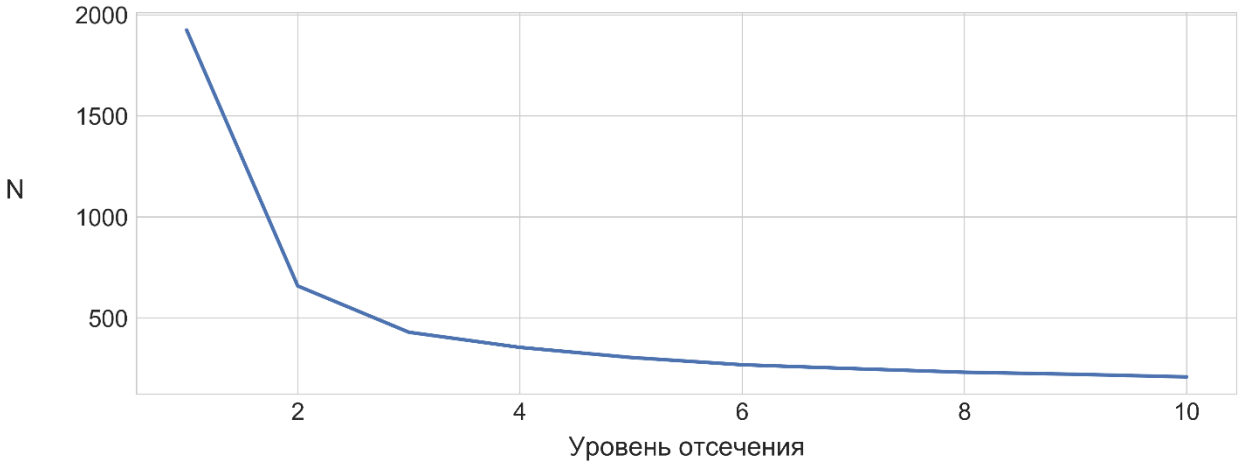

Рис. 2 Количество уникальных имён в зависимости от носителя

Как видим, признак Имя почти у трети объектов обучающей выборки можно закодировать лишь 10 новыми бинарными, однако это приведёт к сильной потери информации, и как следствие, малой обобщающей способности, построенной на таких данных, модели. С другой стороны, в случае применения, к примеру, One-Hot кодирования для абсолютно всех имён (1266 из которых встречаются лишь один раз), велика вероятность не только впустую потратить вычислительные ресурсы, но и получить множество признаков с нулём для большинства объектов обучающей выборки, что может также негативно сказаться на модели. Аналогичная ситуация наблюдается и у другого номинального признака с большим числом уникальных значений (Приложение, Рисунок 1).

Информация о проведенных сменах, на основе которой необходимо получить целевую переменную, в сыром виде содержит порядка 30 различных признаков, однако большинство из них носят технический характер (к примеру, информация о менеджере смены или подтверждение оплаты), поэтому, основываясь на консультациях с представителем работодателя, были выбраны основные количественные и качественные показатели за смену, информация о которых представлена в Таблица 2:

Параметр	Тип	Количество	Unique	Top
QTotalCalcType	Бинарный	78197	2	По выработке
Статус смены (Смена)	Номинальный	78197	11	Подтвержден
Явка на смене (Смена)	Бинарный	78197	2	Да
Тип биллинга	Номинальный	78197	2	Первичный
QTotal	Количественный	64572	-	-

Таблица 2. Смены

2.2 Предобработка

По мере развития алгоритмов и методов машинного обучения всё острее встаёт вопрос предварительной обработки данных, так как даже самая лучшая модель на зашумлённых и противоречивых данных вряд ли покажет результат лучше, чем просто случайный выбор возможного ответа. Отмечается, что сегодня порядка 60–80% аналитической работы в сфере ML — подготовка и анализ данных.

2.2.1 Бинаризация целевой переменной

Явно целевая переменная в исходных данных не представлена, поэтому её требуется синтезировать самостоятельно. Реализована функция [2], осуществляющая преобразование всей информации о смене в множество $\{0, 1\}$ с учётом заданного уровня отсечения по норме выработки.

2.2.2 Обработка пропусков

Обработка пропущенных значений — важный этап подготовки данных. С одной стороны, наличие пропущенных значений является одним из факторов, снижающих достоверность получаемых результатов, с другой стороны, неверная процедура заполнения пропущенных значений может внести шум и противоречивость в модель. Однако одно неизменно: исключить из данных пропуски тем или иным образом необходимо, так как большинство алгоритмов машинного обучения в явном виде часто не могут их обработать.

Возвращаясь к Таблице 1. Признаки, можно заметить, что данные имеют достаточно плотную и заполненную структуру, это объясняется прежде всего тем, что они изначально были подвержены обработке пропущенных/некорректных значений на этапе экспорта из CRM.

Однако несколько признаков по-прежнему имеют пропуски, и их количество не позволяет применить тривиальное исключение всей записи, где встречается пропущенное значение. Наименее заполненной является информация об электронной почте, но при консультации с работодателем выяснилось, что данный признак не является обязательным (именно этим

можно объяснить его сильную разреженность), поэтому все пропущенные значения заменяются на специальное значение *«Не указано»*. Далее происходит очистка и приведение к общему виду имеющихся значений, к примеру:

$$\{yindex, ya, yndex\} \rightarrow yandex,$$

а все непопулярные домены электронной почты (чаще всего это личный домен) заменяются на специальное значение *«Other»*.

Принимая во внимание факт того, что признак Город, где пропуски отсутствуют, обладает иерархической зависимостью[4] от признака Субъект федерации, реализована возможность получения региона из информации о городе (используется самостоятельно полученный словарь). Записи с пропущенными значениями в признаке Возраст, для которых имеется признак Дата рождения, тривиально восстанавливаются, иначе запись удаляется полностью.

Признак	**Количество**		**Unique**	
	-	FillNA	-	FillNA
ID (автономер в базе)	38937	38647	-	-
Дата рождения	38647	38647	11356	11356
Возраст	38635	38647	-	-
Субъект федерации	36582	38647	69	72
E-mail	15363	38647	217	10

Таблица 3. До/после заполнения пропусков

Результатом этапа предобработки стало незначительное снижение объёма на 1% от исходного, но при этом обеспечивается исключительная корректность восстановленных значений.

[4] Отношение, характеризующееся подчинением нижестоящего вышестоящему в данной иерархической системе.

2.2.3 Преобразование и получение новых признаков

Стоит отметить, что признаки формата *datetime* в явном виде могут не только не улучшить модель, но и внести в данные зашумлённость, что вовсе приведёт к падению качества работы. В исходных данных таким признаком является Дата рождения, процесс преобразования которого схематично представлен ниже:

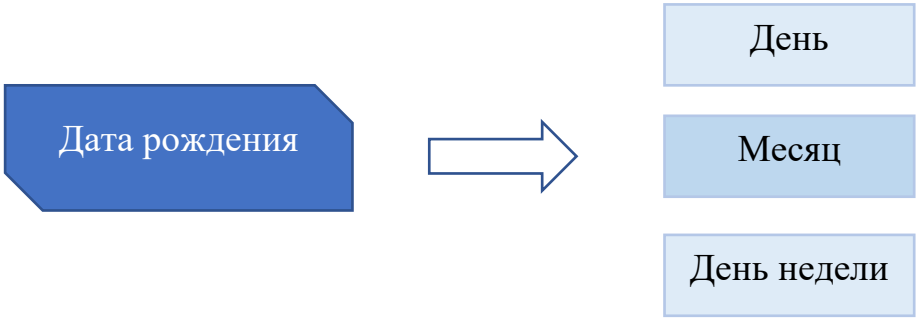

Рис. 3

Хотя с научной точки зрения знак зодиака никак не должен влиять на характер и продуктивность человека, однако, согласно последним опросам [26], почти половина респондентов отмечают между ними корреляцию. Поэтому, учитывая отмеченную ранее близость темы данной работы к психологии и социологии, решено рассмотреть вариант включения в итоговую модель признака, содержащего знак зодиака человека.

Следующим признаком, который также требуется видоизменить по причине его малой информативности в явном виде, является номер мобильного телефона (в исходных данных имеется информация о первых пяти цифрах).

Предлагается два подхода: суть первого в получении 3 цифр кода телефона вида 9** (с заменой наименее популярных на специальное значение), второго – преобразование кода в название оператора, для этого используются данные из открытых источников, в спорных ситуациях отдаётся предпочтение тому оператору, у которого больше номеров с таким кодом.

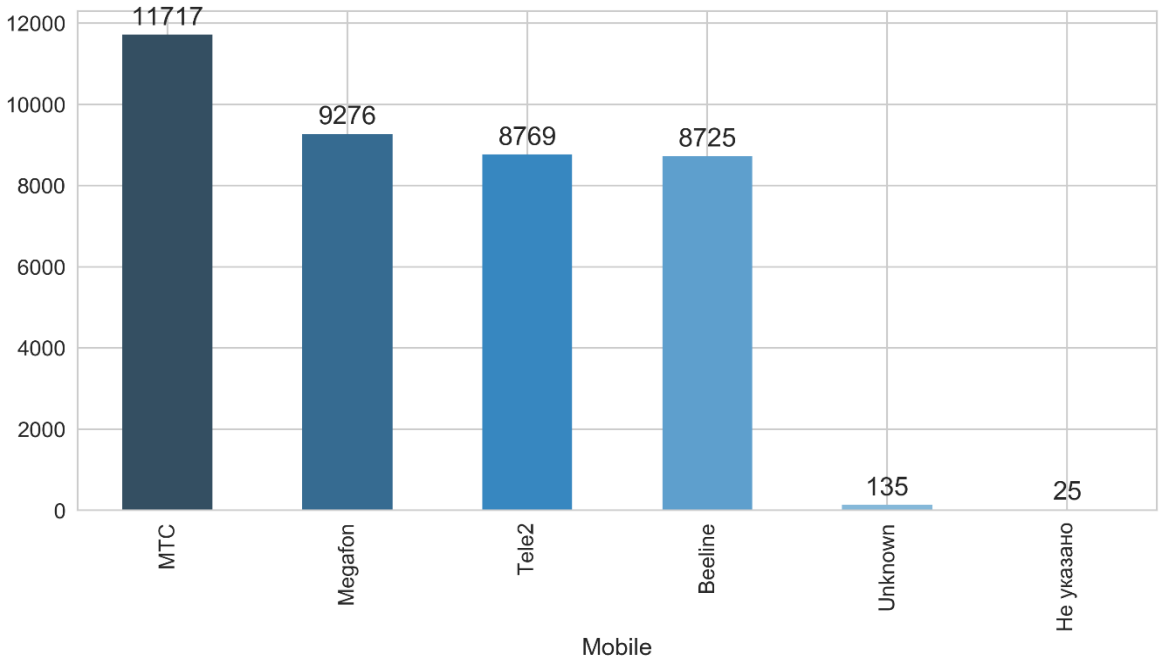

Рис. 4 Статистика по операторам

Итогом преобразования номера телефона в название оператора является понижение на несколько порядков количества уникальных значений данного признака, получившееся распределение по операторам представлено на графике выше.

2.2.4 Категориальные признаки

Категориальным (номинальным/факторным) признаком называется признак, значение которого в той или иной мере обозначает принадлежность объекта к некоторой категории, или же, по другой формулировке, наличие у объекта определенного свойства.

Исходные данные изобилуют подобными признаками, однако, как отмечалось ранее, ни один из алгоритмов машинного обучения явно с ними работать не способен (по причине того, что методы классификации/регрессии формулируются в терминах евклидовых пространств), поэтому все номинальные признаки необходимо преобразовать в числовое значение. Чаще всего выделяют следующие методы кодирования:

Label Encoding

Сопоставление некоторого числа каждой категории: значение категориального признака заменяется на число в соответствии со словарём. Очевидным недостатком является сильное упрощение данных, а также

введение ложных операций сравнения и интерпретации (неявно вводится алгебра над категориями). Однако, существуют методы, для которых этот недостаток не играет никакой роли, к примеру, Random Forest. С другой стороны, явным преимуществом является сохранение размерности данных, что в случае с большим числом уникальных значений у категориального признака, может стать существенным. Возможная реализация *LabelEncoder* для всех категориальных признаков предложена в [2].

OneHot (Dummy) Encoding

Альтернативный и самый простой метод, лишенный озвученного выше недостатка, когда для каждого категориального признака создаются N новых бинарных признаков (N – число уникальных значений исходного признака). К сожалению, простота и естественность подобного кодирования привели к очевидному недостатку: сильное *раздувание*, то есть увеличение размерности, матрицы признаков, и, как следствие, невозможность применения на больших объемах данных без значительного улучшения вычислительного ресурса, поэтому *Dummy*-кодирование не рекомендуется применять, к примеру, для решающих деревьев и бустинга над деревьями, так как обилие различных признаков приводит к существенному увеличению размеров деревьев.

Оба метода, рассмотренные выше, имеют особенность: множество возможных категорий предполагается статичным; и если с точки зрения теории это можно назвать даже скорее плюсом, то для практического применения это налагает существенное ограничение – невозможность категориальному признаку принимать новые значения, отличные от тех, что были изначально. В качестве возможного решения данной особенности был предложен основанный на эвристике метод хеширования признаков [14]. Другой способ — добавление каждому категориальному признаку нового сигнального значения *New Feature* и кодирование неизвестных категорий им, именно второй способ применяется в текущей работе.

Глава III. Сравнение различных классификаторов

В данной главе приводятся результаты применения различных методов машинного обучения к полученным по итогам предыдущей главы данным.

3.1 Методика промежуточного тестирования

С учётом необходимости объективного тестирования различных алгоритмов классификации набор данных был разделён на два непересекающихся множества: тренировочное и тестовое в соотношении 7 к 3 — на тренировочную часть отводится 70% записей, и оставшиеся 30% – на тестовое множество; с целью подбора оптимальных гиперпараметров классификаторов используется скользящий контроль по 10 фолдам (fold).

С одной стороны, исходя из экономических соображений в контексте задачи, для оценки качества классификации следует выбрать точность (*precision),* однако исследуя и максимизируя только точность можно получить слишком низкую полноту (*recall*) классификации, что в среднесрочной перспективе может повлечь уменьшение лояльности возможных сотрудников к *stuffing* агентству. Поэтому в качестве основной меры качества классификации выбрана площадь под ROC-кривой (*англ.* area under ROC curve): как по причине того, что решаемая задача лишена явного дисбаланса классов (в таком случае следовало бы выбрать PR-кривую), так и по причине того, что ROC AUC является одним из стандартов оценивания качества бинарной классификации, при этом исследуется как количественная характеристика, так и сама кривая.

Для большинства задач, связанных тем или иным образом с экономикой и возможными финансовыми потерями от неверно принятого *положительного* решения, приоритетным является снижение доли FP (см. Таблица 4), и если, к примеру, на банковском рынке понижение полноты сказывается некритично для продолжения дальнейшей банковской деятельности, то на рынке временного персонала с его большой текучестью

кадров, слишком низкая полнота классификатора может привести к невозможности осуществления принятых на себя обязанностей по причине нехватки кадров. Поэтому в качестве одного из критериев к ожидаемому решению был нижний порог по полноте на уровне 0,3.

		Реальный класс	
		Positive	*Negative*
Предсказанный класс	*Positive*	TP	FP
	Negative	FN	TN

Таблица 4

3.2 Программное обеспечение

При выполнении данной работы использовалось следующее программное обеспечение:

- Microsoft Excel, Trifacta – первичный анализ и получение общего представления о структуре предоставленных данных;
- PyCharm

Функции предобработки, модификации исходных данных, а также сравнение различных классификаторов реализованы с использованием языка программирования Python (версия 3.6), в качестве дополнительных пакетов использовались:

- Pandas – open-source библиотека, для анализа и обработки данных
- NumPy – open-source библиотека, для работы с матрицами
- Matplotlib – библиотека для визуализации двумерной графики
- scikit-learn – open-source библиотека с реализацией различных алгоритмов машинного обучения
- seaborn – библиотека для визуализации графики

3.3 Классификаторы

3.3.1 Логистическая регрессия

В качестве первого базового алгоритма выбрана широко применяемая в эконометрике и статистическом анализе логистическая регрессия (далее *LR*), которая часто используется для решения задач бинарной классификации. Преимущества и недостатки *LR*:

+	устойчивость	–	большой объём данных
+	малое время обучения	–	сложность отбора признаков
+	апостериорная вероятность	–	чувствительность к масштабу признаков
+	приемлемое качество		

Для *LR* доступен выбор следующих гиперпараметров:

- L_1 или L_2 норма регуляризации[5] – *penalty*
- C – величина обратная коэффициенту регуляризации
- Алгоритмы настройки весов – *solver*
- Веса классов: равные или сбалансированные.

Поиск оптимальных гиперпараметров осуществлён при помощи реализованного в *sklearn* перебора по сетке – *GridSearchCV* – суммарно 48 различных комбинаций параметров, лучшие результаты представлены ниже.

Ранг	ROC AUC Test	ROC AUC Train	C	Weights	penalty	solver
1	**0,745830894**	0,775416378	1	None	l1	liblinear
2	0,745721847	0,775554292	1	balanced	l1	liblinear
3	0,74498064	0,783835312	1	None	l2	liblinear
3	0,74498064	0,783835312	1	None	l1	liblinear
5	0,744935587	0,783743592	1	None	l1	lbfgs
6	0,744917326	0,783792936	1	balanced	l2	liblinear
6	0,744917326	0,783792936	1	balanced	l1	liblinear

Таблица 5. GridSearchCV для LR по всем признакам

[5] L_2 — при мультиколлинеарности признаков, L_1 — при необходимости отбора признаков

Заметим, что у всех классификаторов, находящихся на вершине списка, коэффициент регуляризации равен единице, что в совокупности с преобладанием L_1–регуляризации может косвенно свидетельствовать об излишнем количестве признаков; отсутствие преобладания баланса весов говорит о том, что оба класса в исследуемых данных уже имеют исходно сбалансированную структуру. Для проверки гипотезы об излишнем количестве признаков с помощью метода *SelectFromModel* были отобраны наиболее важные, и вновь запущен *GridSearchCV* – суммарно 16 различных комбинаций параметров, результаты подтвердили исходное предположение:

Ранг	ROC AUC Test	ROC AUC Train	C	Weights	penalty	solver
1	**0,751562982**	0,776504058	10	None	l2	liblinear
2	0,751439845	0,776342932	10	balanced	l2	liblinear
3	0,751419575	0,77678489	10	None	l1	liblinear
3	0,751298717	0,776933467	100	None	l2	liblinear
5	0,751280299	0,776599188	10	balanced	l1	liblinear
6	0,751273023	0,776946558	100	None	l1	liblinear
6	0,751161406	0,776735595	100	balanced	l2	liblinear

Таблица 6. GridSearchCV для LR с отбором признаков

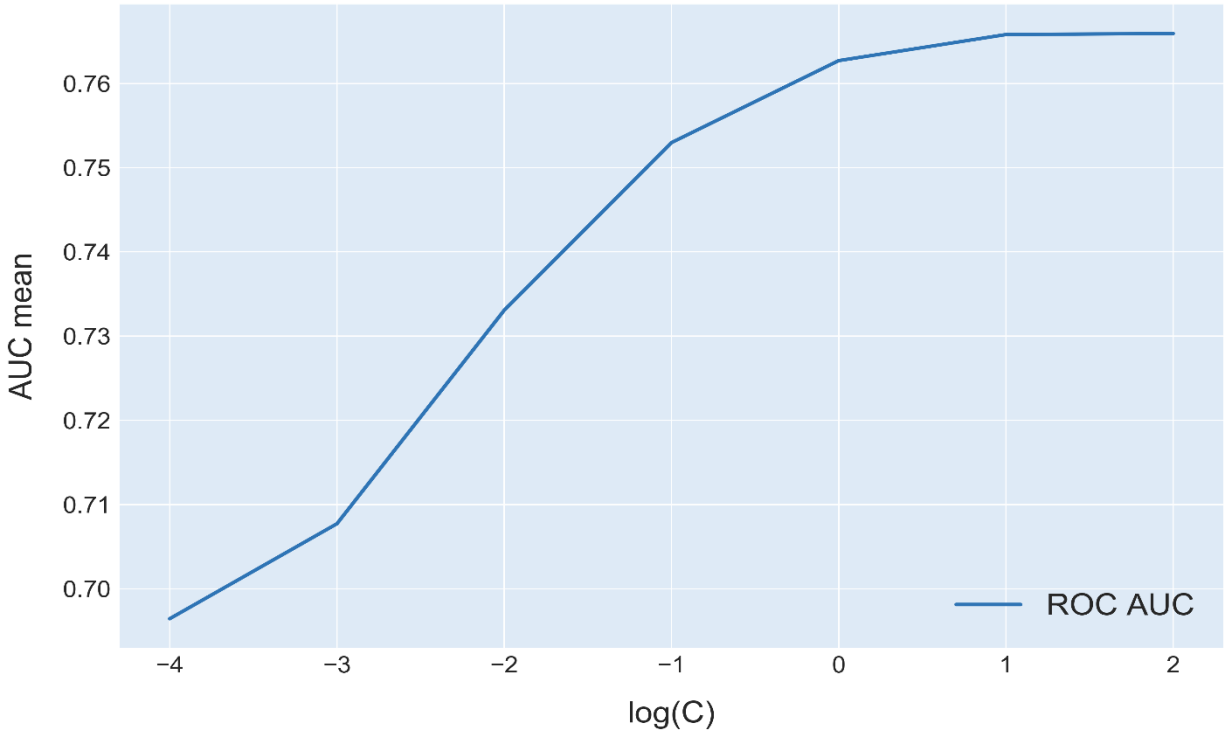

Рис. 5 Зависимость ROC AUC от параметра C

На Рис. 5 показан рост площади под ROC-кривой при росте параметра C на кросс-валидации, то есть наблюдается обратная зависимость от коэффициента регуляризации; график ROC–кривой, построенной для лучшего классификатора LR на отдельном тестовом множестве приведён ниже:

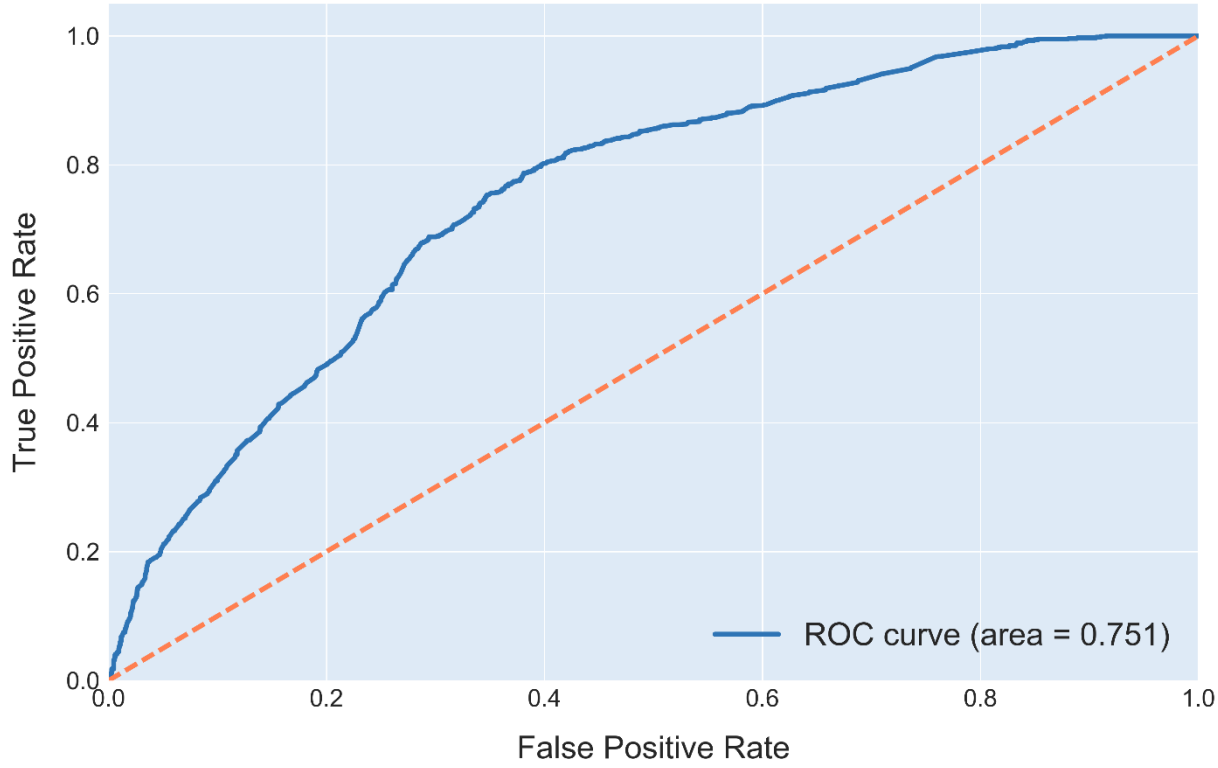

Рис. 6 Результат лучшего LR на тестовом множестве

3.3.2 Нейронная сеть

Искусственная нейронная сеть (ANN) — один из самых известных методов машинного обучения, однако лишь с ростом вычислительных мощностей нейронная сеть (далее *Neural*) обрела популярность как практический инструмент. Среди преимуществ и недостатков *Neural* выделяют:

+	высокая точность	–	длительность обучения
+	обобщающая способность	–	*Black Box*
+	нелинейность	–	большой объём данных
+	устойчивость	–	требовательность к Hardware

Для класса *MLPClassifier*, реализующего возможности *Neural* в *scikit*,

предлагается выбор следующих параметров:

- Функция активации для скрытых слоёв
- α – коэффициент регуляризации
- Алгоритмы настройки весов – *solver*
- Максимальное число итераций
- Число нейронов на скрытых слоях[6]

Поиск оптимальных гиперпараметров осуществлён по сетке – *GridSearchCV* – в общей сложности 60 различных комбинаций, лучшие результаты представлены ниже.

Ранг	ROC AUC Test	ROC AUC Train	α	Итераций	Структура	solver
1	**0,745099076**	0,780750104	0,01	100	(100,)	logistic
2	0,745099076	0,780750104	0,01	200	(100,)	logistic
3	0,744173089	0,927751287	0,01	100	(100,)	relu
3	0,744173089	0,927751287	0,01	200	(100,)	relu
5	0,742995385	0,787616319	0,01	100	(50, 50)	logistic
6	0,742995385	0,787616319	0,01	200	(50, 50)	logistic
6	0,74298596	0,788095738	0,01	100	(30, 20, 5)	logistic

Таблица 7. GridSearchCV для Neural по всем признакам

Коэффициент регуляризации у лучших методов оказался величиной постоянной (как и в случае с *LR*), однако теперь он направлен на возможное увеличение весовых коэффициентов, и, как следствие, поощрение более сложной границы между классами, что может быть связано с недообучением (*underfitting*) прежде всего из-за специфики таких признаков как Имя, Отчество (см. 2.1 Анализ). Выделяется несущественное влияние изменения структуры сети на результат: *TOP3* представлен исключительно структурой по умолчанию; малозначимым оказалось и изменение числа итераций – свидетельство того, что после 100 итераций сходимость незначительна. Проверим предположение, озвученное выше, путём фильтрации категориальных признаков Имя, Отчество – реализована

[6] Кортеж, где *i*-й элемент означает число нейронов на *i*-м скрытом слое.

специальная функция (подробнее [2]) по замене редких имён и отчеств на специальное значение «Редкое» – и последующим повторным запуском *GridSearchCV* – суммарно 8 различных комбинаций параметров:

Ранг	ROC AUC Test	ROC AUC Train	α	Итераций	Структура	solver
1	**0,753701687**	0,806182656	1	100	(50, 50)	relu
2	0,745082512	0,766185438	1	100	(100,)	relu
3	0,743990385	0,773473063	0,01	100	(100,)	logistic
3	0,742580654	0,928474748	0,01	100	(100,)	relu
5	0,742399334	0,778746074	0,01	100	(50, 50)	logistic
6	0,720826771	0,93167739	0,01	100	(50, 50)	relu
6	0,709759288	0,714158458	1	100	(100,)	logistic

Таблица 8. GridSearchCV с фильтрацией Имени и Отчества

Действительно, согласно результатам, представленным в Таблица 8, сокращение числа уникальных значений у двух категориальных признаков (порог отсечения для имени и отчества – 2 и 1, соответственно) позволило повысить результат.

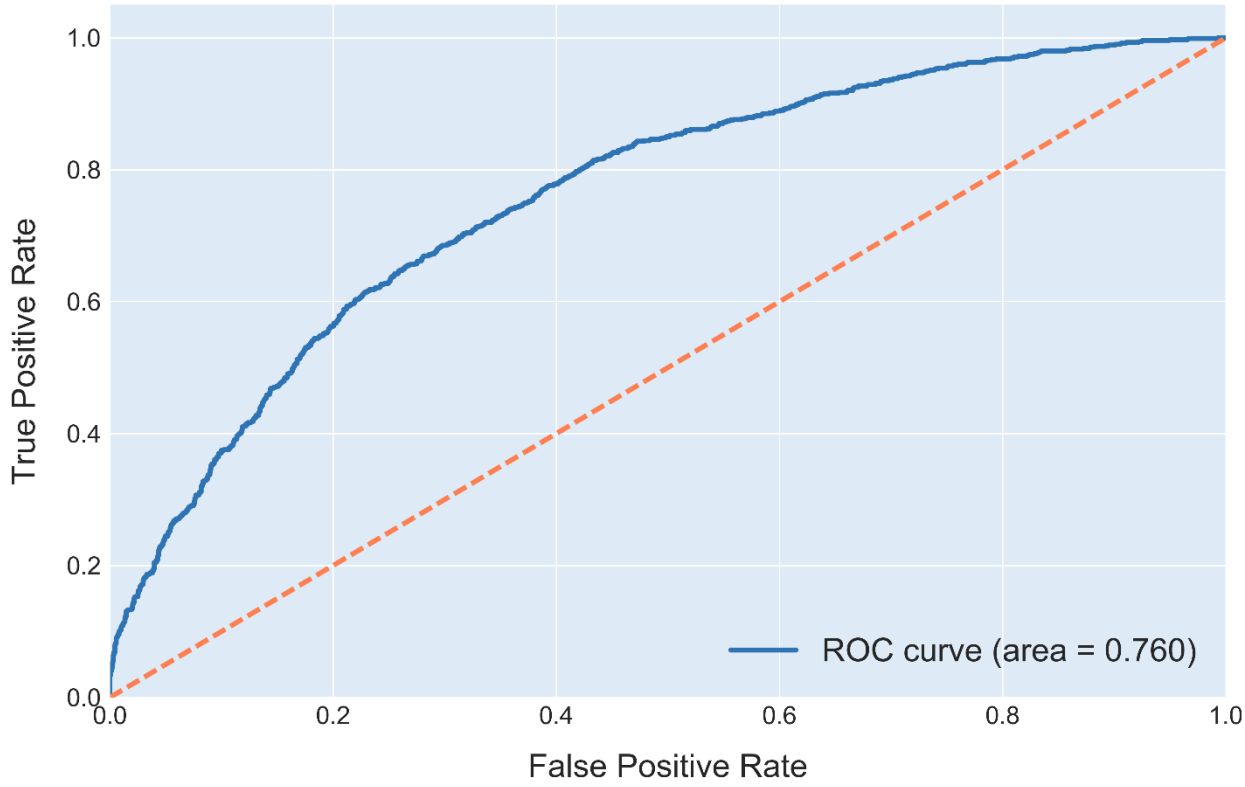

Рис. 7 Результат лучшего Neural на тестовом множестве

Дальнейший перебор установил, что оптимальным с точки зрения

максимизации качества является уровень отсечения равным 2 для обоих признаков; график ROC–кривой, построенной для этого классификатора на отдельном тестовом множестве приведён на Рис. 7.

3.3.3 Случайный лес

В современном машинном обучении редко можно встретить методы, которые не были бы существенно улучшены с момента их открытия, однако случайный лес (англ. *Random Forest*, далее *RF*) по-прежнему применяется в своём *первозданном* виде – так, как он был разработан на рубеже тысячелетий. Главные достоинства и недостатки этого метода перечислены ниже:

+ высокая масштабируемость	– размер построенной модели
+ эффективность работы с большим числом признаков	– требовательность к размеру обучающей выборки
+ нечувствительность к масштабу признаков	– время обучения и работы

Отдельно отметим, что, во-первых, в отличие от двух рассмотренных ранее методов, *RF* работает с *Label Encoding*, и во-вторых, *RF* позволяет оценивать важность признаков. В качестве гиперпараметров для *RF* доступны:

- Число деревьев – при увеличении растет качество, однако пропорционально увеличивается и время обучения
- Критерий расщепления: энтропийный или Джинни
- Число признаков, участвующих в выборе лучшего расщепления
- Максимальная глубина

Для подбора оптимальных гиперпараметров вновь воспользуемся *GridSearchCV* – 96 различных комбинаций, наибольшая вариативность по числу деревьев и максимальному числу признаков, которые участвуют в определении лучшего расщепления на каждой итерации; параметры, показавшие лучшие результаты, перечислены в Таблица 9.

Ранг	ROC AUC Test	ROC AUC Train	Глубина	Число деревьев	Число признаков[7]	Критерий расщепления
1	**0,758937755**	0,85420324	10	1000	0,5	gini
2	0,758621878	0,846187525	10	1000	0,5	entropy
3	0,758374391	0,8539865	10	200	0,5	gini
4	0,758359642	0,84567899	10	200	0,5	entropy
5	0,758300335	0,853464203	10	500	0,5	gini
6	0,758201922	0,845283827	10	500	0,5	entropy
7	0,758059262	0,85336471	10	100	0,5	gini

Таблица 9. GridSearchCV для RF

Во-первых, сразу отмечается превосходство всех лучших классификаторов *RF* над базовыми вариантами *LR* и *Neural*, однако повышение качества далось ценой увеличения как времени построения модели, так и времени получения результата. Во-вторых, результаты практически инвариантны критерию расщепления, что часто говорит об отсутствии явной специфичности деревьев, входящих в лес, но, учитывая контекст задачи, более вероятным является предположение о зашумленности данных или же малой важности многих признаков, за счёт чего различия в реализации критериев несущественны. Воспользуемся особенностью *RF* – оценим важность признаков, и проверим:

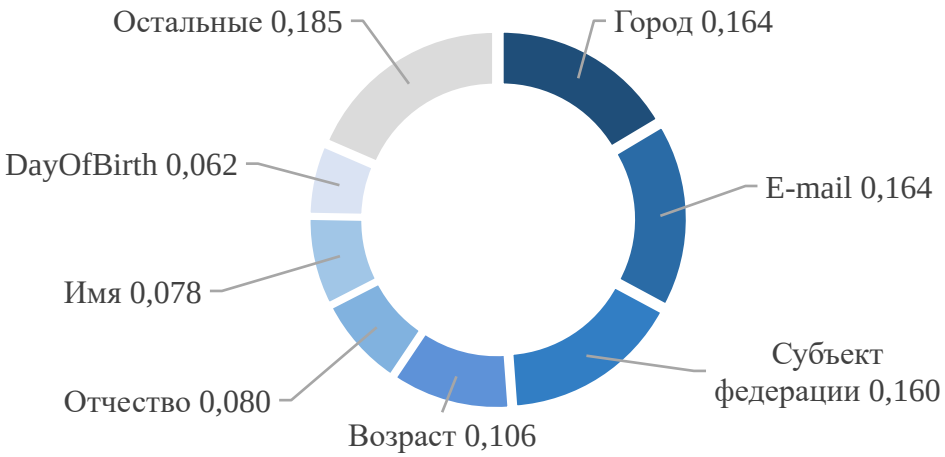

Рис. 8 Распределение важности признаков

Анализ круговой диаграммы, представленной выше, действительно позволяет утверждать, что данные, по которым строится *RF*, сильно зашумлены как

[7] Вещественное число означает процент от общего числа признаков

минимум по таким признакам как Имя и Отчество: *RF* склонен считать признаки с наибольшим числом уникальных значений как более важные, что сейчас далеко от истины.

Для увеличения качества классификации были предприняты следующие шаги:

- ✓ замена непопулярных значений признаков Имя, Отчество

В отличие от *Neural*, пороги отсечения носят значительно более радикальный характер: к примеру, для Отчества оптимальная граница, найденная перебором, составляет 200, а для имени – 250, что в целом удовлетворяет концепции *RF*.

- ✓ отбор оптимального подмножества признаков

Базовое множество признаков составляют те, важность которых более 10%, согласно Распределение важности признаков. Различные комбинации дополнительных признаков и соответствующие им результаты представлены ниже:

Имя	Отчество	Zodiac	Mobile	ROC AUC Test
	+	+	+	**0,762247700**
	+			0,760770155
	+	+		0,760575713
	+		+	0,758801752
		+	+	0,758574254
+	+			0,758525860
			+	0,757352293
+				0,757252911
		+		0,756980908
+		+		0,756769831
				0,753731779

Таблица 10. Отбор дополнительных признаков

Из таблицы выше можно заключить, что признак Имя наименее важен, а часто и вовсе лишь ухудшает качество классификации; добавление признака Отчество, напротив, всегда улучшает результат классификации. Отдельно

отмечаем синергию таких признаков как *Zodiac* и *Mobile*, раздельное добавление которых к Отчеству приводит к ухудшению модели, и лишь именно их совместное включение позволяет улучшить результат.

Итог, удалось как повысить качество модели (Рис. 10), так и снизить время обучения и последующей классификации, при этом важность обработанных признаков, напротив, возросла (следует заметить, что рост не в абсолютных значениях, а относительных: общее количество уникальных значений признака снизилось на два порядка, а вклад остался прежним):

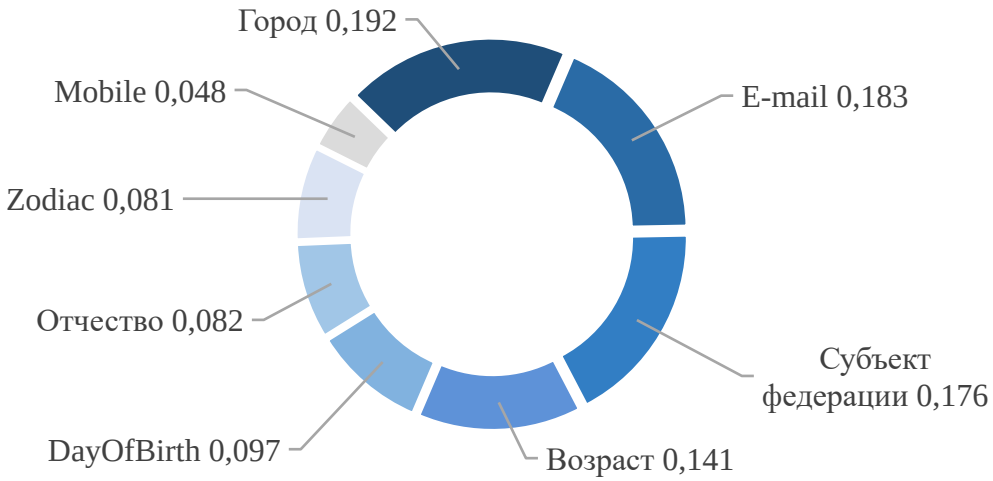

Рис. 9 Распределение важности признаков после обработки

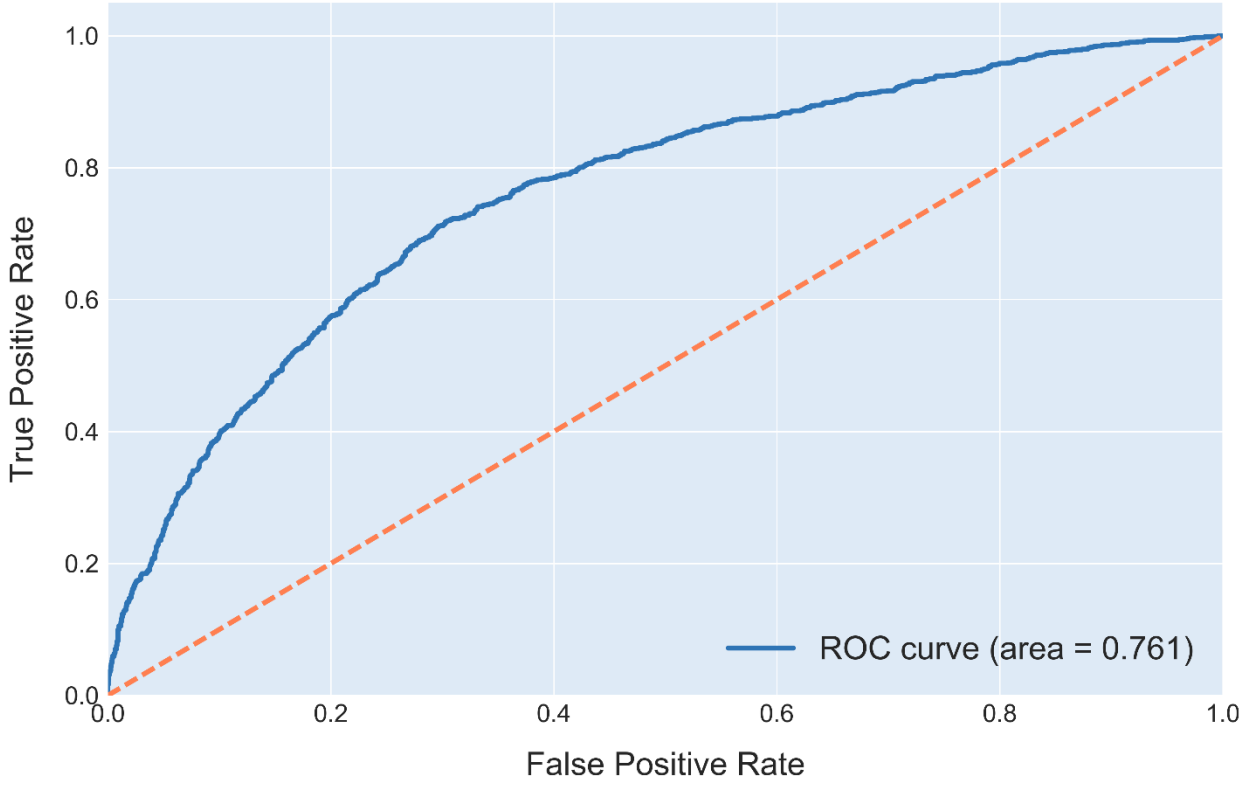

Рис. 10 Результат лучшего RF на тестовом множестве

3.4 Анализ результатов

Предыдущие пункты данной работы позволили получить 3 лучших классификатора, различных по природе построения модели; как было отмечено в 3.1 Методика промежуточного тестирования, для поиска оптимального классификатора использовалась ROC AUC – мера качества, которая действительно позволяет объективно оценивать качество классификации *в целом*, то есть без явного смещения классификатора к минимизации либо ошибок I, либо ошибок II рода, однако она скорее *синтетическая*. В контексте рассматриваемой задачи нас интересует точность, и при этом имеется нижняя граница по полноте, поэтому для всех трёх классификаторов на тестовом множестве был построен график PR-кривой: лучший результат при условии *Recall* > 0.3 показал классификатор, относящийся к группе *RF*, график представлен ниже:

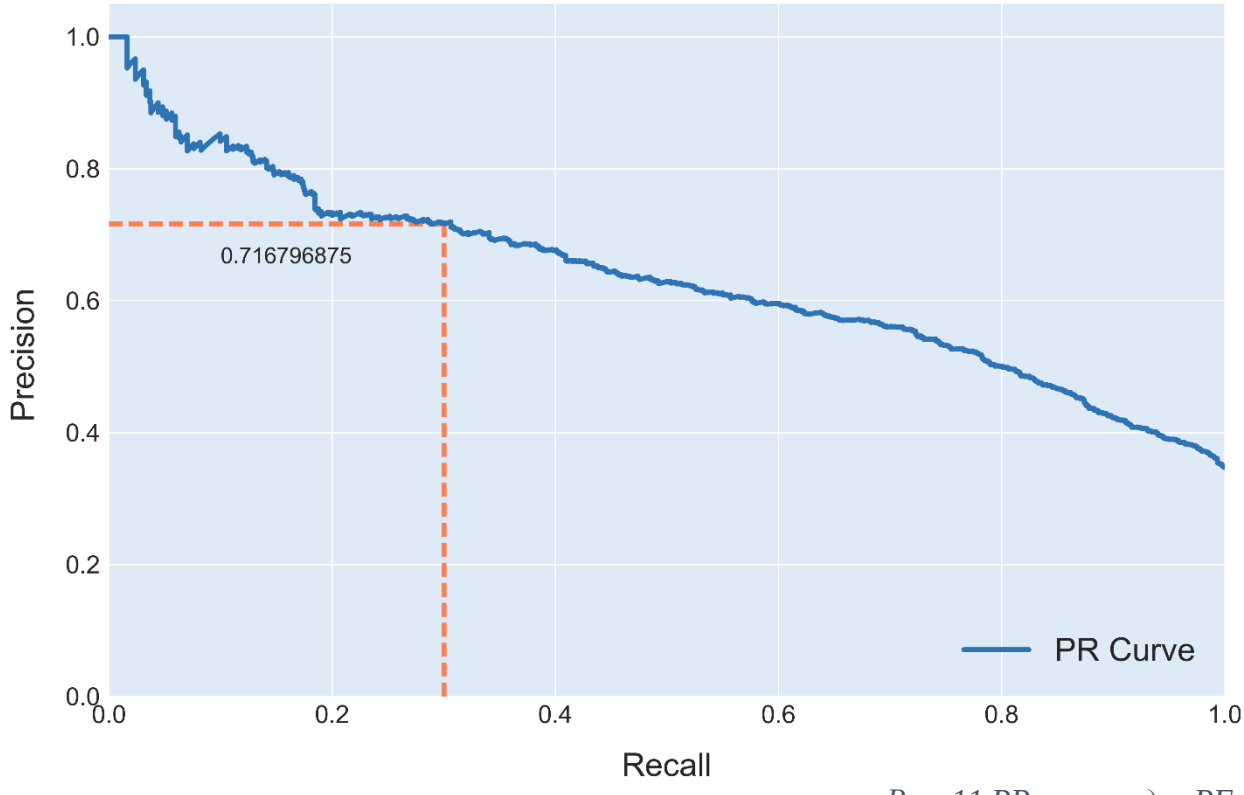

Рис. 11 PR-кривая для RF

Самую низкую точность, обеспечивающую полноту выше 30% продемонстрировал классификатор группы *LR* – 0.624, *Neural* же расположился между ними – 0.675 (графики приведены в Приложении).

С одной стороны, достигнутый результат характеризуется как хороший, и даже в чистом виде может быть внедрён в существующие бизнес-приложения *stuffing* агентств. С другой стороны, вызывает недоумение значительное отставание по точности классификатора *LR*, который по значению ROC AUC почти не уступает двум другим. Была выдвинута гипотеза, что *LR* лучше подходит для определения отрицательного класса, и именно за счёт этого значение площади под ROC-кривой высоко, отчасти на это указывает и более равномерное снижение PR-кривой (Приложение, Рисунок 3) нежели у конкурирующих *Neural* и *RF*.

Поэтому было решено на основе полученных моделей построить ансамбль, который, согласно результатам базовых лучших классификаторов, строит финальное предсказание. В качестве информации, получаемой от базовых классификаторов, выступает вероятность принадлежности положительному классу. К сожалению, средняя вероятность базовых вероятностей улучшила результаты на тестовом множестве совсем незначительно (относительно *RF*); но перебором были найдены веса базовых классификаторов, максимизирующие значение точности ансамбля:

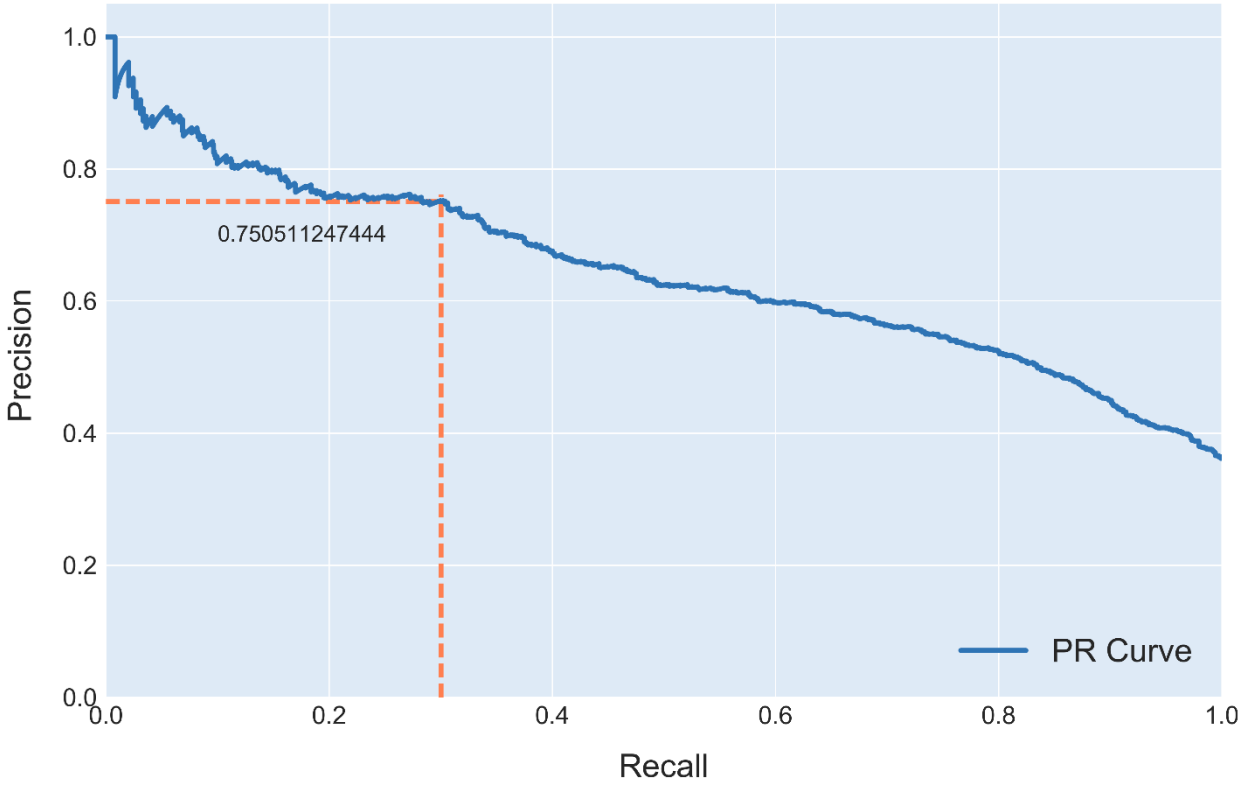

Рис. 12 PR-кривая для ансамбля базовых классификаторов

Выводы

В данной работе была исследована одна из возможностей интеграции искусственного интеллекта в бизнес-процессы HR-агентств, специализирующихся на рынке временного персонала.

Для выполнения исследования прежде всего было необходимо определить существующие сферы соприкосновения *AI* и *HR*, контекст этого соприкосновения, его плюсы и минусы, а также глубже ознакомиться с предметной областью по причине её особой специфики: близости как к психологии, так и к экономике, что, как следствие, накладывает дополнительные требования к ожидаемому решению.

В ходе работы осуществлён анализ исходных данных, результаты которого позволили определить наиболее перспективные методы их предобработки и модификации, произведено объективное сравнение моделей, построенных на различных методах машинного обучения, продемонстрированы варианты их оптимизации. Результатом стала программная реализация интеллектуальной системы, которая по входным данным – персональной информации потенциального сотрудника – делает прогноз его *качества*, при этом интеллектуальная система удовлетворяет всем выдвигаемым требованиям, среди которых максимизация точности предсказания с заданной нижней границей полноты и перспектива адаптации к смежным сферам деятельности.

Таким образом, был продемонстрирован вариант приложения искусственного интеллекта к рынку подбора кадров, поэтому поставленную цель можно считать достигнутой.

Предложенный вариант имеет несколько путей дальнейшего усовершенствования, среди которых:

- с точки зрения *Data*:
  - включение в число признаков информации частного характера: к примеру, банковской для стратификации по финансовым

показателям, или личных предпочтений в искусстве для формирования психологического портрета;

- кластеризация категориальных признаков Имя, Отчество в зависимости от результата сотрудника за смену и замена их явного использования на кластер, к которому они принадлежат;
  - с точки зрения *ML*:
- определение характера человека, опираясь на его социальную активность (к примеру, при помощи LIWC анализа), и включение полученного результата в качестве признака;
- применение Deep Learning.

Заключение

По итогам исследования достигнуты следующие результаты:

+ рассмотрены присутствующие на рынке варианты применения машинного обучения в контексте *HR*;
+ на основе анализа исходных данных, определены оптимальные методы их предварительной обработки и модификации;
+ построены различные модели классификации, осуществлено их объективное сравнение;
+ предложена реализация интеллектуальной системы, интеграция которой в бизнес-процессы одного из *stuffing* агентств России намечена на лето 2017.

Приложение

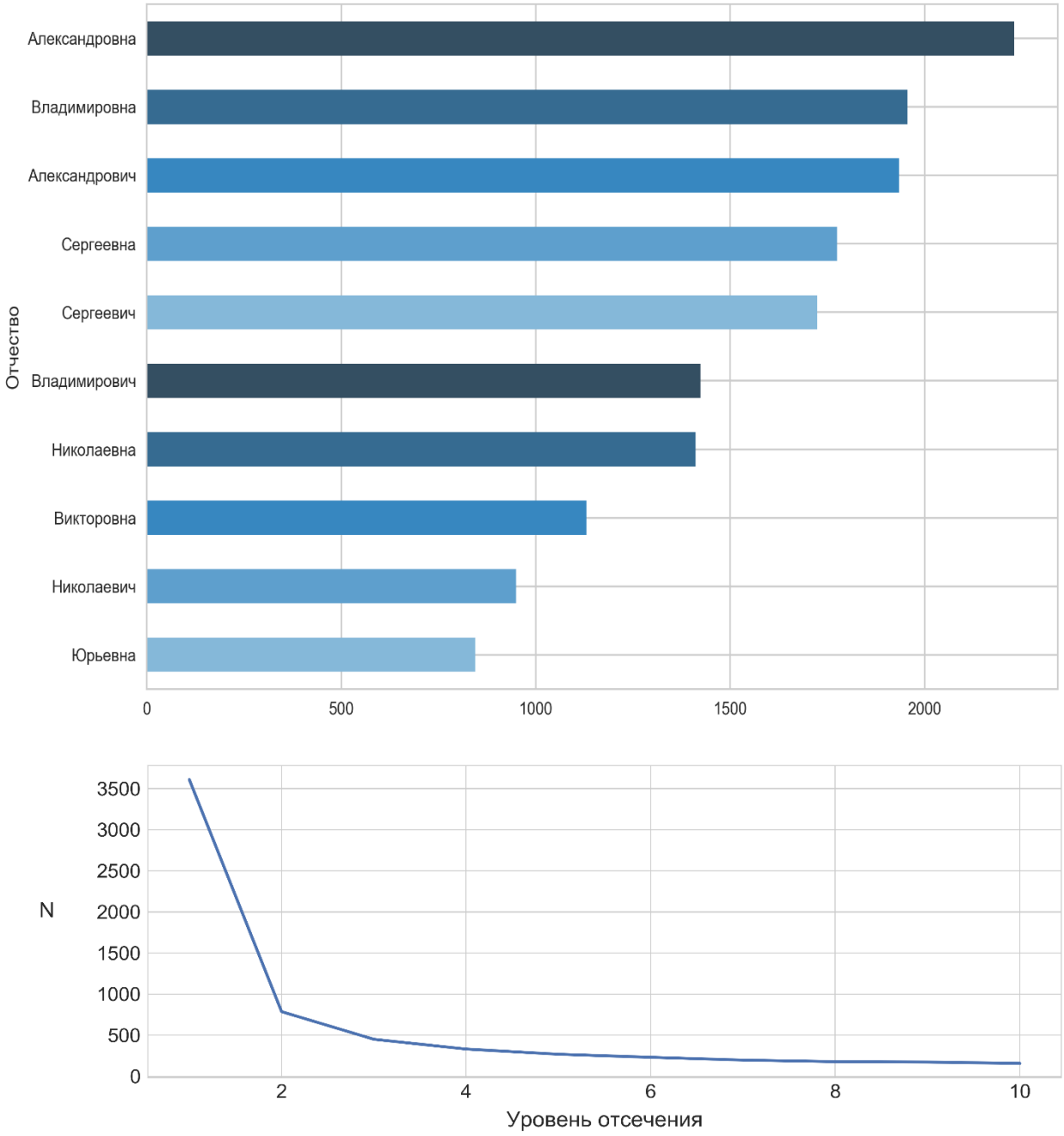

Рисунок 1. Признак Отчество

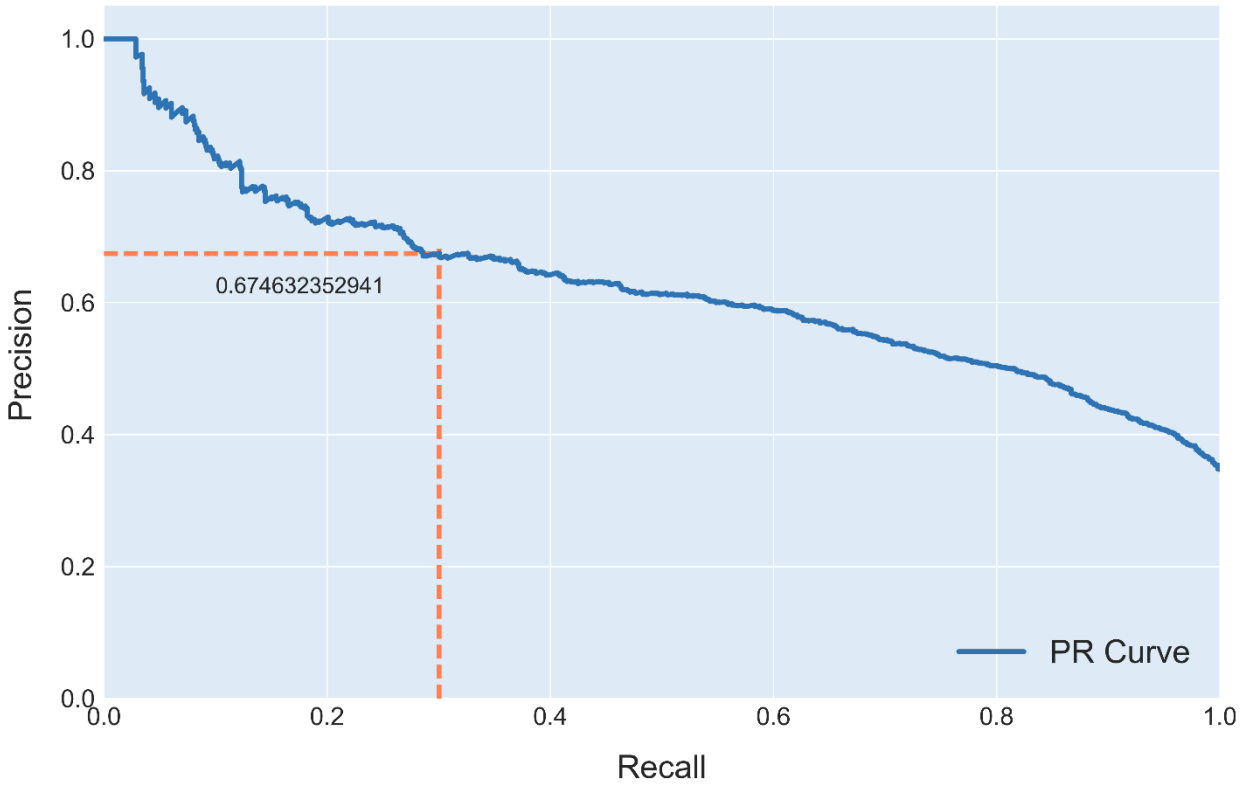

Рисунок 2. PR-кривая для Neural

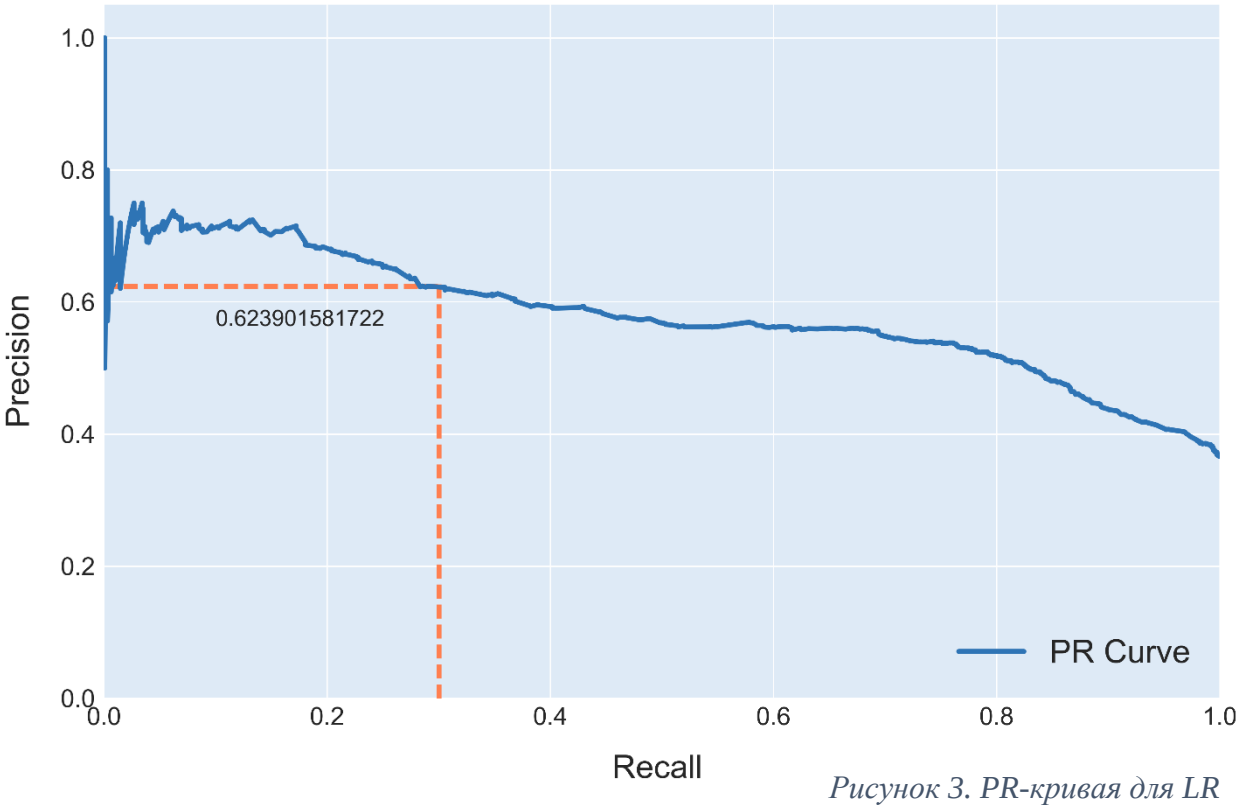

Рисунок 3. PR-кривая для LR

Список литературы

1. ABBYY. Обработка данных из резюме [Электронный ресурс] // abbyy.com: [сайт]. URL: https://www.abbyy.com/ru-ru/solutions/resume/
2. AIRY [Электронный ресурс] // github.com: [сайт]. URL: https://github.com/AlexWorldD/AIRY
3. DataScience. Machine learning - features engineering from date/time data [Электронный ресурс] // datascience.stackexchange.com: [сайт]. URL: https://datascience.stackexchange.com/questions/2368/machine-learning-features-engineering-from-date-time-data
4. Faliagka E., and Tsakalidis A., and Tzimas G. An integrated e-recruitment system for automated personality mining and applicant ranking // Internet research. 2012. Vol. 22. No. 5. pp. 551-568.
5. IBM. IBM Kenexa Talent Acquisition Suite [Электронный ресурс] // ibm.com: [сайт]. URL: https://www-03.ibm.com/software/products/ru/talent-acquisition
6. Kulkarni R. How to Leverage AI for Talent Management [Электронный ресурс] // HRTechnologist.com: [сайт]. URL: https://www.hrtechnologist.com/articles/productivity-analysis-hr-analytics-tools/how-to-leverage-ai-for-talent-management/
7. LinkedIn. Will This Year’s College Grads Job-Hop More Than Previous Grads? [Электронный ресурс] // linkedin.com: [сайт]. URL: https://blog.linkedin.com/2016/04/12/will-this-year_s-college-grads-job-hop-more-than-previous-grads
8. Pande S. E-recruitment creates order out of chaos at SAT telecom: system cuts costs and improves efficiency // Human Resource Management International Digest. 2011. Vol. 19. No. 3. pp. 21-23.

9. Peoplise. We digitize your recruitment process and empower HR teams with smart use of technology and data. [Электронный ресурс] // peoplise.com: [сайт]. URL: http://www.peoplise.com/
10. Pravda.ru. Временный персонал: как оптимизировать процесс поиска, найма и администрирования [Электронный ресурс] // pravda.ru: [сайт]. URL: https://www.pravda.ru/navigator/podbor-vremennogo-personala.html
11. re:Work. Let's Make Work Better. [Электронный ресурс] // rework.withgoogle.com: [сайт]. URL: https://rework.withgoogle.com/
12. Rice M.E., Harris G.T. Comparing Effect Sizes in Follow-Up Studies:ROC Area, Cohen’s d,andr // Law and Human Behavior. 2005. No. 5. pp. 615-620.
13. Ulrich D. Ulrich's new dawn for HR October 2012. pp. 41-43.
14. Weinberger K., and Dasgupta A., and Langford J., and Smola A., and Attenberg J. Proceedings of the 26th Annual International Conference on Machine Learning // Feature hashing for large scale multitask learning. 2009. pp. 1113-1120.
15. Zimmermann T., and Kotschenreuther L., and Schmidt K. Data-driven HR-Resume Analysis Based on Natural Language Processing and Machine Learning // arXiv preprint arXiv:1606.05611, 2016.
16. Аникин Б.А., Рудая И.Л. Аутсорсинг и аутстаффинг: высокие технологии менеджмента. 2-е изд. М.: ИНФРА-М, 2009. 287-290 с.
17. Аутсорсинг [Электронный ресурс] // ru.wikipedia.org: [сайт]. URL: https://ru.wikipedia.org/wiki/Аутсорсинг
18. Дворская Е. Робот, зарплату повыше [Электронный ресурс] // forbes.ru: [сайт]. URL: http://www.forbes.ru/tehnologii/331897-robot-zarplatu-povyshe-kak-mashinnoe-obuchenie-i-novye-tehnologii-raboty-s-dannymi

19. Зинченко А.А. Массовый подбор персонала в энергетической отрасли с использованием математических методов // Экономика и предпринимательство. 2015. № 7. С. 794-798.
20. Зинченко А.А. Применение нейросетевых моделей для принятия решений о подборе персонала // Вестник Тамбовского университета. Серия: Естественные и технические науки. 2015. Т. 20. № 2.
21. Качаева Т.В., Южиков В.С. Автоматизированная система распознавания и классификации резюме // В кн.: Труды российской конференции молодых ученых по информационному поиску в рамках RuSSIR 2007. Екатеринбург. 2007. С. 64-72.
22. Салогуб А.М., Демина Н.В. Новые тенденции в управлении талантливым персоналом и HR-технологий // Гуманизация образования. 2015. № 2. С. 105-113.
23. Сафронов А.В. Resumagic: система автоматической обработки резюме.
24. Тагиров В.К., Тагирова Л.Ф. Интеллектуальная поддержка принятия решений в задачах подбора персонала на основе композиционных правил нечеткой логики // Технические науки – от теории к практике. 2013. № 4. С. 19.
25. Титова С.В. Методика оценки резюме кандидатов на вакантную должность в организации // Наука Красноярья. 2016. Т. 3. № 2. С. 106-112.
26. Фонд "Общественное мнение". Об отношении к астрологии и гороскопам [Электронный ресурс] // http://fom.ru: [сайт]. URL: http://fom.ru/obshchestvo/11435
27. Шутина О.В., Реут И.Ю. Особенности управления временным персоналом // Вестник Омского университета. Серия «Экономика». 2014. № 4. С. 68-75.